RECOMMENDATION SYSTEM

BY,

LUBNA RAHMAN

INTRODUCTION

- Recommendation system is an information filtering system that seeks to predict the rating or the preference a user might give to an item.
- In other words, it is an algorithm that suggests relevant items to users.
- Ex: which movie to watch in Netflix, in e-commerce which product to buy, or in kindle which book to read, and many more.

TYPES OF RECOMMENDATION SYSTEM

i. Content Based Filtering

ii. Collaborative Based Filtering: is of two types as mentioned below,

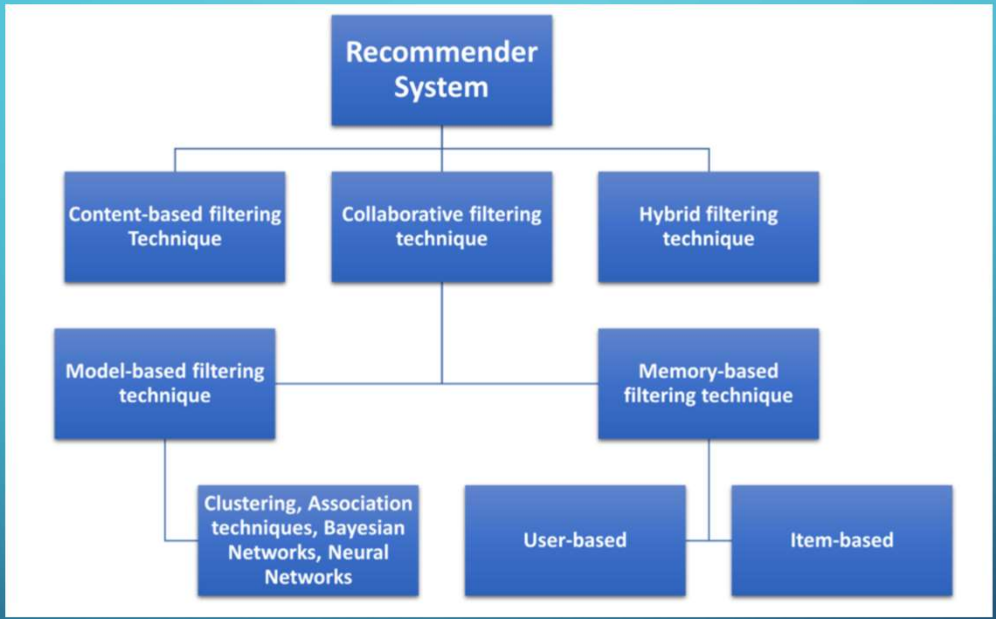
a) Memory based filtering: is again divided into two subtypes ;

- User Based Collaborative Filtering
- Item Based Collaborative Filtering

b) Model based filtering

iii) Hybrid Filtering : (Content based+ Collaborative based approach)

TYPES OF RECOMMENDATION SYSTEM



COLLABORATIVE VS CONTENT

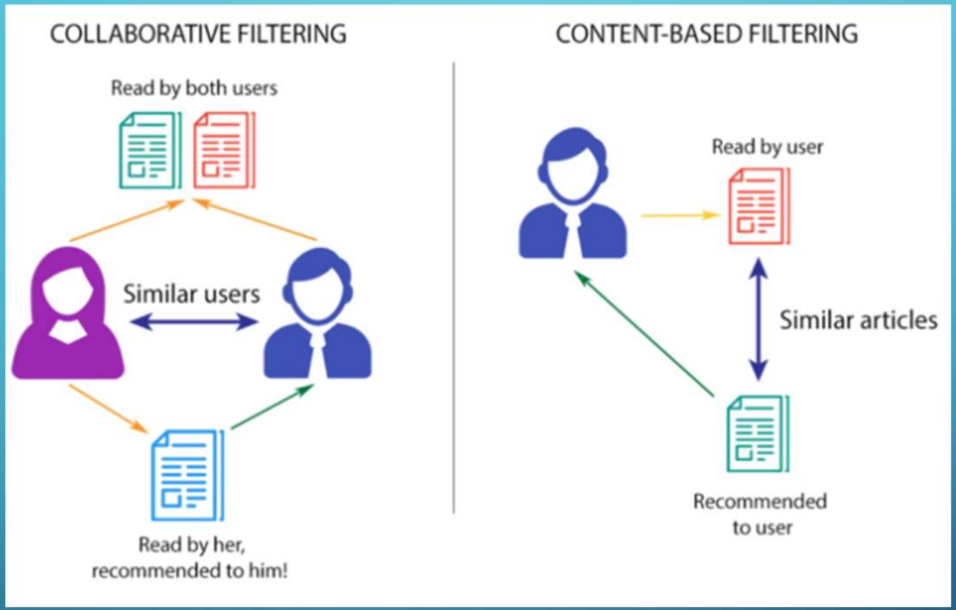


Fig. 1.1

CONTENT BASED FILTERING

- Content-based filtering uses item features to recommend other items similar to what the user likes, based on their previous actions or explicit feedback.
- In other words, relevant items are recommended using the content of the previously searched items by the users.
- The content or attributes of the things you like are referred to as "content.“ For ex. Genre of the movie.

ADVANTAGES OF CONTENT BASED FILTERING

- Easy to Scale: Because the recommendations are tailored to a person, the model does not require any information about other users. This makes scaling of a big number of people more simple.
- Specific: The model can recognize a user's individual preferences and make recommendations for things that only a few other users are interested in.
- Faster: New items may be suggested before being rated by a large number of users.

DISADVANTAGES

- <u>Requires Domain Expertise</u>: This methodology necessitates a great deal of domain knowledge because the feature representation of the items is hand-engineered to some extent.
- <u>Limited potential</u>: The model can only give suggestions based on the user's current interests. In other words, the model's potential to build on the users' existing interests is limited.
- <u>Lack of novelty</u>: Since it must align the features of a user's profile with available products, content-based filtering offers only a small amount of novelty.

COLLABORATIVE BASED FILTERING

- Collaborative based filtering is recommending the new items to users based on the interest and preference of other similar users. For instance, Amazon recommending new products saying "Customer who brought this also brought".
- It has two sub-types as mentioned earlier:

  1. Memory based filtering
  2. Model based filtering

COLLABORATIVE BASED FILTERING

MEMORY-BASED COLLABORATIVE FILTERING

Memory based collaborative filtering has two subtypes:

a) Used-based collaborative filtering:

Rating of the item is done using the rating of neighboring users. In simple words, it is based on the notion of users' similarity. Similar users will be recommended similar items. For ex: recommending similar movies to similar users (users who give similar ratings to the same movies)

b) Item-based collaborative filtering: In this case, first similarities between items are computed then based on the computed similarities, items similar to already consumed/rated are looked at and recommended accordingly. In simple words, it is based on the notion of item similarity.

USER BASED FILTERING VS ITEM BASED FILTERING

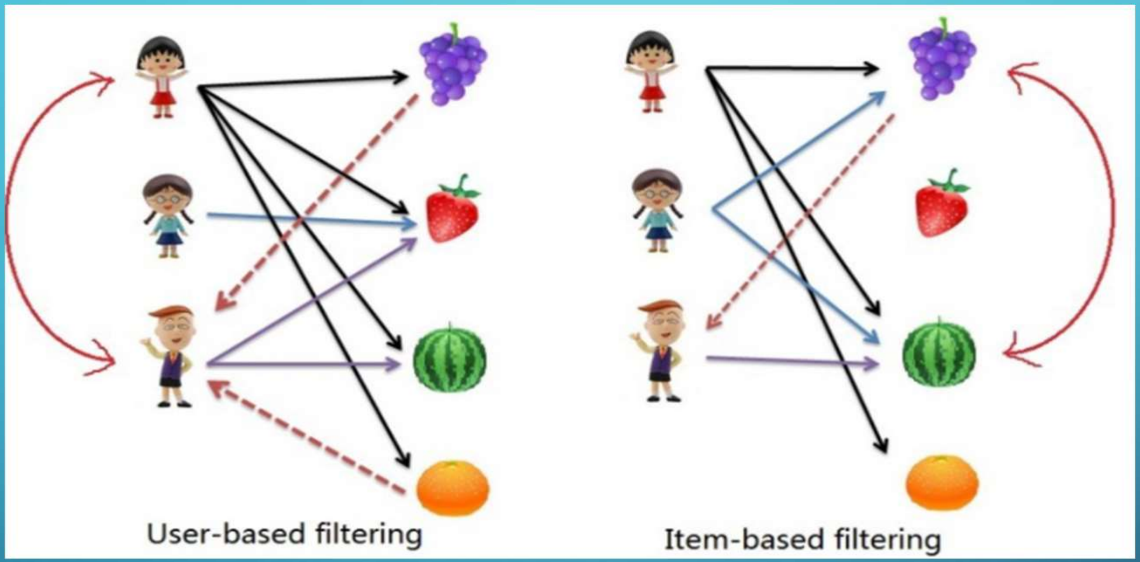


Fig. 1.2

COLLABORATIVE BASED FILTERING

MODEL BASED COLLABORATIVE FILTERING

Model based collaborative filtering also called as Latent Factor Model(LFM)

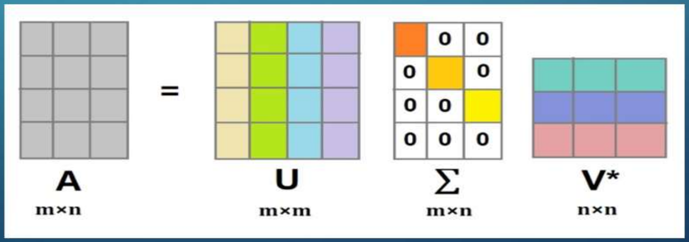
- Latent factor model factorize user ratings of products into user and item feature vectors. By taking this approach, the model aims to find low-rank feature matrices to describe the data.
- It is one of the most successful methods for Collaborative filtering in the recommendation system.
- Latent factor models such as Singular Value Decomposition (SVD) extract features and correlation from the user-item matrix.

COLLABORATIVE BASED FILTERING

SINGULAR VALUE DECOMPOSITION(SVD)

- Uses dimensionality reduction technique to find latent factors or the hidden features.
- Breaks down a matrix into a product of a few smaller matrices.
- Reduces a high rank matrix to a lower rank matrix while preserving the important information.
- Rank of a matrix is a measure of the unique information stored in the matrix. Higher the rank, more the information.

Fig. 1.3



APPLICATIONS OF SVD

- Image compression: reducing the size of the image.
- Spectral clustering: gives a perfectly clustered data.
- Eigenfaces: human face recognition.
- Page ranking: used by search engines like Google to select which page should be displayed first in a search.
- Correlation matrix creation: shows the relationship between different types of users and items.

ADVANTAGES OF COLLABORATIVE FILTERING

- It works well even if the data is small.
- This model helps the users to discover a new interest in a given item as the model recommends it because similar users are interested in that item.
- No need for domain knowledge.

DISADVANTAGES OF COLLABORATIVE FILTERING

- It cannot handle new items because the model doesn't get trained on the newly added items in the database. This problem is known as Cold Start Problem.
- Side feature doesn't have much importance. Ex: Here side features can be actor name or releasing year in the context of movie recommendation.

EVALUATION METRICS

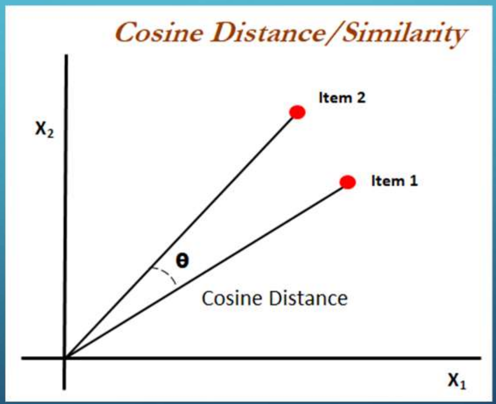
Evaluation metrics are used to evaluate whether the given model is recommending the right things or not and how many relevant things this system predicts. Some of the popular evaluation metrics are:

- Cosine Similarity
- Mean Average Recall at K
- Coverage
- Personalization
- MSE (Mean Square Error) and RMSE(Root Mean Square Error)
- Intra-list similarity
- Confusion Matrix

EVALUATION METRICS
COSINE SIMILARITY

Cosine Similarity: It is a metric used to measure how similar two items are by measuring the cosine of the angle between two vectors/items projected in a multi-dimensional space. The output value ranges from 0–1. A value of 0 means no similarity and 1 means both the items are 100% similar.

Fig. 1.4



EVALUATION METRICS

- Mean Average Recall at K: It shows how much relevant is the list of recommended items. Here precision at K means the recommended items in top k sets that are relevant.
- Coverage: It is the percentage of items in the training data model able to recommend in test sets. Or simply, the percentage of a possible recommendation system can predict.
- Personalization: It shows how many same items, the model recommends to different users i.e. the dissimilarity (1-cosine similarity) between users lists and recommendations and a great way to assess if a model recommends many of the same items to different users.

EVALUATION METRICS

- MSE and RSME: measure the average of the squares of the errors, i.e. the average squared difference between the estimated values and the actual values. The lower the error, better the model.
- Intra-list similarity: It is the average cosine similarity of all items in a list of recommendations. This calculation uses features of the recommended items (such as movie genre) to calculate the similarity.
- Correlation matrix: It is a table showing the correlation coefficients between sets of variables . In other words, in a recommendation system shows the relationship between the users and the items of the dataset.

WHY ARE RECOMMENDATION SYSTEM NECESSARY

- The long tail phenomenon makes recommendation systems necessary.
- The distinction between the physical and online worlds has been called the long tail phenomenon, as displayed in the figure below :

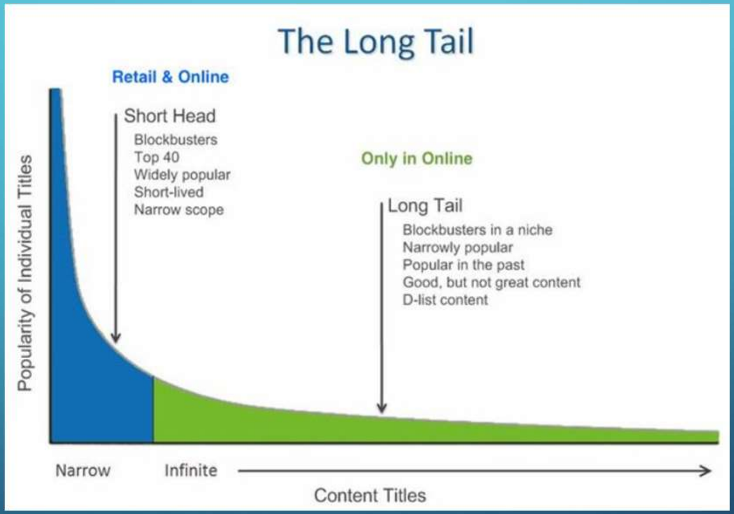


Fig. 1.5

THE LONG TAIL PHENOMENON

- The vertical axis in the figure represents the popularity of items and the horizontal axis, the options based on their popularity.
- Physical stores can only offer popular products, to the left of the vertical line, on-line stores on the other hand, can provide the entire range of items, the tail as well as the popular.
- The role of the recommendation system is to suggest items to users that include popular ones and those on the long tail.
- The long-tail phenomenon forces online institutions to recommend items to individual users.

APPLICATIONS OF RECOMMENDATION SYSTEM

- e-commerce: product recommendation
- Movie/songs recommendation
- Books recommendation
- Job recommendation
- Online courses recommendation
- Restaurant recommendation
- New house/rental property recommendation
- Places to visit/things to do in a city recommendation

BENEFITS OF RECOMMENDATION SYSTEM

- Increased sales/conversion: can help generate a huge amount of income when they are efficient and can be a way to stand out significantly from competitors.
- Increased user satisfaction: It reduces customers path to a sale by recommending them an appropriate option, sometimes even before they search for it.
- Increased loyalty and share of mind: getting customers to spend more on the website, it can increase their familiarity with the brand and user interface, increasing their probability to make future purchases.
- Reduced churn: Recommendation system-powered emails are one of the best ways to re-engage customers. Discounts or coupons are other effective yet costly ways of re-engaging customers and they can be coupled with recommendations to increase customers' probability of conversion.

QUESTIONS?

THANK YOU!